Sleeping Beauty in Quantumland

Lev Vaidman

November 13, 2001

Abstract

It is argued that thirder resolution of the Lewis - Elga controversy about Sleeping Beauty is more clear when the coin toss is replaced by a quantum measurement and the analysis is performed in the framework of the Many-Worlds Interpretation.

1 Lewis - Elga controversy

Lewis's comment (2001) on Elga's paper (2000) has far reaching consequences. If Lewis is right, then the approach to credence for an event as the value of an "intelligent bet" on this event (e.g. Sklar 1993) does not have universal applicability. The betting approach to this question (Aumann et al. 1997) leads to Elga's result which Lewis contests. I believe, however, that Lewis's approach is untenable, and thus the universality of the betting approach to probability has not been breached.

I give two arguments in favor of Elga's conclusion, the first classical, the second quantum mechanical. It is of interest that a quantum mechanical analysis can be brought to bear on this classical probability problem.

The first is as follows: consider the following three experiments, the first of which is Elga's. I claim the credence of Beauty is the same in each case.

> (i) Beauty sleeps during the week. On Sunday a fair coin is tossed. If the result is Heads (H), Beauty is awaken on Monday only, if the result is Tails (T), Beauty is awaken on Monday and Tuesday. On every awakening, Beauty is told nothing, but must answer the question: "What is your credence for the coin to be

> H?" After the conversation her memory of the awakening is erased and she sleeps again.
>
> (ii) Beauty sleeps for 100 years, uninterrupted save for awakenings according to procedure (i), which take place every week, following a coin toss on each Sunday of each week.
>
> (iii) Beauty sleeps for 100 years, uninterrupted save for 7827 awakenings. Using a classical random number generator to determine the order, on 2609 awakenings a coin is placed H, and on 5218 awakenings a coin is placed T. On each awakening, the same procedure is followed as in (i).

Experiment (ii) is not just a repetition of experiment (i) 5218 times. In the latter, on every awakening, Beauty knows which week it is. In (ii) Beauty does not know which week it is and, therefore, she does not know which coin in the sequence the question is about. However, since all the coins are fair, Beauty's situation is the same in all weeks and thus the lack of information which week it is cannot make a difference in her credence.

Experiments (ii) and (iii) are also not identical. The probability in experiment (ii) that one will obtain exactly 2609 H will is very small. However, the probability that this number will be different from 2609 by more than, say, 100, is also very small. Therefore, the probability that the relative frequency of H will be significantly different from 1:2 is negligible. Thus, although I do not have an argument according to which Beauty has to give exactly the same answer in (ii) and (iii), I can argue that these answers cannot differ significantly. Since our job is to decide between credences 1/2 (Lewis) and 1/3 (Elga), this is enough.

In experiment (iii) it is obvious that Beauty should give credence 1/3 for H. If there is no significant difference in Beauty's answer in case (iii) and (ii), and no significant difference in her answer in case (ii) and (i), then in Elga's experiment Beauty should give the answer 1/3, and not 1/2 as Lewis claims.

There is a conceptual difference between (i) and (ii) (and a similar difference between (i) and (iii)). In (i) the Beauty was asked a question about an uncentered proposition: "What is the state of the coin?" In contrast, in (ii) Beauty was asked a question about a centered proposition: "What is the state of the coin of *this* week?" The unusual feature of Elga's experiment, that Beauty must alter her credence about an uncentered proposition with no new uncentered evidence, is not present here. But this does not alter the fact that Beauty must give the same answer in case (ii) and (i).

Apart from the small statistical difference between (ii) and (iii) already mentioned, the two differ in another aspect. In (ii) there is something which corresponds to Lewis's number 1/2: in half of the weeks of Beauty's sleep the coin is H, and in half it is T. In contrast, in experiment (iii), nothing corresponds to 1/2. However, knowledge of this statistical structure to her string of awakenings, in case (ii), is not knowledge that Beauty can use, since never on awakening does she learn where in the string she is located.

1.1 The inconsistency of the Lewis approach

In order to see how the difference between the one-week experiment and the many-weeks experiment arises, and in order to show the inconsistency of Lewis's approach, consider an experiment of the kind (ii) but limited to two weeks. A fair coin is tossed twice. The credence of Beauty $p(H)$ can be calculated as the sum of conditional probabilities on the outcomes of the coin tosses:

$$p(H) = p(H|HH)p(HH)+p(H|TT)p(TT)+p(H|HT)p(HT)+p(H|TH)p(TH) \tag{1}$$

It is uncontroversial that $p(H|HH) = 1$ and $p(H|TT) = 0$. Given that one of the outcomes is H and another T, Beauty knows that there are three awakenings: one H and two T. Therefore, the conditional credences of Beauty for these cases are $p(H|HT) = p(H|TH) = \frac{1}{3}$. According to Lewis, Beauty has equal credence for all possible outcomes of the coin tosses: $p(HH) = p(TT) = p(HT) = p(TH) = \frac{1}{4}$. It then follows from (1) that Beauty's credence for H on awakening during the two weeks is $p(H) = \frac{5}{12}$. This is in contradiction with the assumption that there should be no change between Elga's one-week experiment and the similar two-week experiment. Therefore, unless Lewis rejects this very natural assumption, his approach is inconsistent.

On the analysis that I favor there is no such difficulty. On awakening, Beauty's credences for the four outcomes of the coins tosses should not be identical, they should be weighted according to the number of awakening corresponding to these outcomes. Thus $p(HH) = \frac{1}{6}$, $p(TT) = \frac{1}{3}$, and $p(HT) = p(TH) = \frac{1}{4}$. Using (1), it follows that $p(H) = \frac{1}{3}$, just as in the one-week experiment in accordance with Elga's argument.

2 The quantum coin experiment

An entirely independent argument to the same conclusion follows from the interpretation of probability in the Many-Worlds Interpretation (MWI) of quantum mechanics (an elaboration of the Everett approach (1957)). Consider the toss of a quantum coin, say, for a simple example, the observation of a photon, incident on semi-transparent mirror, as either reflected (R) or transmitted (T). According to the MWI, the world splits in two: one (the R-world) in which the photon is observed as reflected, and the other (the T world) in which the photon is observed as transmitted.

Elga's experiment, but with such a quantum coin, is very similar to the "sleeping pill" experiment (Vaidman 1998), which was introduced to give a possibility of an ignorance interpretation of probability in the framework of the MWI. According to this approach, the observer assigns the probability for outcomes of a quantum measurement in proportion to the "measures of existence" (Vaidman, 1998) of the corresponding worlds, the modulus squares of the amplitudes of the corresponding branches of the universal wave function. (See Saunders (1998) for a discussion of a similar concept of "measure" .) There is no direct meaning for this probability for the person who is going to perform the quantum measurement (and who is put to sleep for its duration): there is no information, centered or uncentered, that he is then ignorant of. The meaning for probability is given through the (identical) credences of the two successors of the experimenter, on awakening, in centered propositions, namely the propositions "I am in the R-world" and "I am in the T-world". For each of these successors, there is a fact of the matter as to the outcome of the experiment and he is ignorant of this fact.

It is worth remarking that the problem of the MWI recently posed by Peter Lewis does not arise in this approach. He argued (Lewis 2000) that a believer in the MWI should agree to play "quantum Russian roulette", provided that death is instantaneous . The large "measures" of the worlds with dead successors is a good reason not to play.

There is a difference between Elga's experiment and the "sleeping pill" quantum coin flipping. On awakening, Beauty is not only ignorant of which world she is in, she is also ignorant of which time she is at in T-world. There are three mutually exclusive propositions: "I am in an H-world on a Monday", "I am in a T-world on a Monday" and "I am in a T-world on a Tuesday", and she must assign credences in them summing to unity. Failing any other information which could discriminate between the three cases, her credences should be in proportion to the "measures" of the corresponding worlds, which happened to be exactly the same. Therefore, her credence in each should be the same, namely one third.

There is an important difference between Elga's experiment with a quantum and a classical coin . Classically, Beauty's credence concerns a proposition, "the coin is H", and this along with the entire sequence of events are located in a unique world. It is for this reason that the proposition is reckoned to be uncentered. But using a quantum coin, on the MWI, that is no longer so. Quantum mechanical outcomes are in different worlds, and propositions about such outcomes, such as "the photon is T" should be read as tacitly indexical, "in *this* world the photon is T"; they are properly speaking centered propositions. To revert to our previous discussion, the difference between case (i) and case (ii), using a quantum coin, does not concern the centeredness or otherwise of the proposition Beauty's credence is about. In both cases it is centered.

On switching to a quantum coin in Elga's experiment, and on interpreting quantum mechanics in terms of the MWI, one loses the unusual feature that Lewis found so objectionable: that Beauty must change her credence in an uncentered proposition, although her uncentered evidence has not changed. Lewis may even *agree*, in this case, that Beauty's credence should change, for it is credence in a centered proposition - and if so he will presumably agree that her credence in H is one third. But the quantum coin is considered to be the best possible implementation of a fair coin: Beauty should not have different credences in classical and quantum experiments.

Acknowledgments

The first part of this work was done together with Simon Saunders (see the discussion of the classical case in Vaidman and Saunders 2001). This research was supported in part by grant 62/01 of the Israel Science Foundation and the EPSRC grant GR/N33058.

References

Aumann, R. J., Hart, S., and Perry, M. 1997 The forgetful passenger, *Games and Economic Behaviour* **20**: 117-120.

Elga, A. 2000 Self-locating belief and the Sleeping Beauty problem *Analysis* **60**:143-147.

Everett, H. 1957 "Relative State' Formulation of Quantum Mechanics," *Review of Modern Physics* **29**: 454-462.

Lewis, D. 2001 Sleeping Beauty: reply to Elga, *Analysis* **61**: 171-176.

Lewis, P. J. 2000 What is it like to be Schrödinger's cat? *Analysis* **60**: 22-29.

Saunders, S. 1998 Time, Quantum Mechanics, and Probability, *Synthese* **114**: 373-404.

Sklar, L. 1993 *Physics and Chance : Philosophical Issues in the Foundations of Statistical Mechanics*, Cambridge : Cambridge University Press, 1993.

Vaidman, L. 1998 On Schizophrenic Experiences of the Neutron or Why We should Believe in the Many-Worlds Interpretation of Quantum Theory, *International Studies in the Philosophy of Science* **12**: 245-261.

Vaidman, L. and Saunders, S. 2001 On Sleeping Beauty Controversy, *PhilSci Archive*, http://philsci-archive.pitt.edu/documents/disk0/00/00/03/24/index.html